

DOCUMENT RESUME

ED 286 936

TM 870 585

AUTHOR Livingston, Samuel A.
TITLE The Effects of Time Limits on the Quality of Student-Written Essays. Revised.
PUB DATE 5 May 87
NOTE 20p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; Difficulty Level; *Essay Tests; Higher Education; High Schools; Holistic Evaluation; *Minimum Competency Testing; *Prewriting; Student Placement; Test Items; *Test Length; *Timed Tests; *Writing Evaluation; Writing Skills
IDENTIFIERS New Jersey College Basic Skills Placement Test

ABSTRACT

The effect of increased writing or planning time on a test of basic college level writing ability was studied. The essay portion of the New Jersey College Basic Skills Placement Test was given to students in nine New Jersey public colleges and three New Jersey public high schools. Each student wrote two essays on two different topics. The first essay allowed 20 minutes writing time. The other allowed either 30 minutes writing time, or 10 minutes planning time plus 20 minutes writing time. There were eight groups altogether, differing on ability, order of topic-writing, and order of longer time allowed. All essays were read by two independent readers and evaluated holistically on a six-point scale. The increased time limit produced very little increase in the students' essay scores, except at high ability levels (i.e., for those students who clearly would not need remedial writing instruction). Adding a 10-minute planning period before the 20-minute writing period tended to increase the scores of the high-ability students and also of the low-ability students who had recently written a 20 minute essay on a similar topic. The largest effect was associated with a difference in difficulty between the two topics used. (Author/JGL)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED286936

The Effects of Time Limits on the Quality
of Student-Written Essays

Samuel A. Livingston
Educational Testing Service

Presented April 24, 1987 at the Annual Meeting of
the American Educational Research Association,
New York, NY.

Revised May 5, 1987

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.
☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Samuel A.
Livingston

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 870 585

ABSTRACT

Increasing the time limit from twenty minutes to thirty minutes on a single-question essay test of basic college-level writing ability produced very little increase in the student's essay scores, except at high ability levels (i.e., for those students who clearly would not need remedial writing instruction). Adding a ten-minute planning period before the twenty-minute writing period tended to increase the scores of the high-ability students and also of the low-ability students who had recently written a 20-minute essay on a similar topic. The largest effect was a difference in difficulty between the two topics used in the study.

ACKNOWLEDGMENTS

The efforts of many people made this study possible. They include the participating students, teachers, and essay readers, the directors of the New Jersey College Basic Skills Placement Test program, and, in particular, Shari Santapau (of the New Jersey Department of Higher Education) and Gertrude Conlan (of Educational Testing Service), who organized and conducted the collection of the data, and Joanne Imrie (of Educational Testing Service), who organized the essay reading sessions.

The New Jersey College Basic Skills Placement Test (NJCBSPT) is actually a battery of five tests in reading, writing, and mathematics skills. All students entering state-supported colleges in New Jersey take the NJCBSPT. The test results are used by the colleges to place students into or out of remedial courses in reading, writing, and mathematics. Each college has its own procedures for placing students; there are no statewide rules for placement.

The essay section of the NJCBSPT consists of a single 20-minute essay. The students are not given a choice of topics and are not given the topic in advance. NJCBSPT essay topics require no special literary or other academic knowledge. The purpose of the essay is to evaluate the students' ability to express their own thoughts in standard written English. The topics are quite general and ask about aspects of the student's personal experience.

The essays are scored on a six-point scale. Each essay is read and scored by two different readers; the student's essay score is the sum of the two scores. The scoring is holistic; readers do not attempt to judge specific aspects of writing quality. The scoring standards are defined by example, using actual student-written essays.

Students taking the NJCBSPT typically write one or two paragraphs in the twenty minutes allowed. Many students' essays are unfinished. This report describes a study conducted to determine the effect of two possible changes in the time limits for the essay. One possible change would be to extend the time limit to 30 minutes. The second possible change would be to add a ten-minute planning period, during which the students are to read

the topic and plan their essays but may not begin writing. In each case, the total time allowed would be increased from twenty minutes to thirty minutes.

Method

The study consisted of two separate experiments, conducted together. Each of the two experiments compared one of the two altered time limits with the original twenty-minute time limit. Each participating student wrote two essays, one essay on each of two different topics. The topics were taken from Forms 3GJP and 3HJP of the NJCBSPT. The order of the time limits and topics was counterbalanced, resulting in the following design:

<u>Group</u>	<u>Time Limit and Topic</u>	
	<u>1st essay</u>	<u>2nd essay</u>
1	20 min.; Topic G	30 min.; Topic H
2	20 min.; Topic H	30 min.; Topic G
3	30 min.; Topic G	20 min.; Topic H
4	30 min.; Topic H	20 min.; Topic G
5	20 min.; Topic G	10 + 20 min.; Topic H
6	20 min.; Topic H	10 + 20 min.; Topic G
7	10 + 20 min.; Topic G	20 min.; Topic H
8	10 + 20 min.; Topic H	20 min.; Topic G

The participants were students in nine New Jersey public colleges and three New Jersey public high schools. They wrote the essays during their regular English classes. Although 676 students wrote essays on Topic G and

626 on Topic H, only 512 students wrote essays on both topics and also provided the identifying information necessary to include their scores in the analysis. For administrative reasons it was impossible to assign individual students randomly to the eight experimental groups. Therefore, classes were randomly assigned to the eight groups. The resulting groups differed in size from 43 to 84 students. Their scores revealed that they also differed somewhat in ability, as can be seen in Table 1.

The essays were administered to each class on two different days, under standardized conditions, from written instructions provided to the participating teachers. The time limit was printed in a statement at the top of the students' instruction sheets. The three versions of the statement were as follows:

Time - 20 minutes. You have twenty minutes to plan and write an essay on the topic assigned.

Time - 30 minutes. You have thirty minutes to plan and write an essay on the topic assigned.

Total time - 30 minutes. You have thirty minutes for this test, ten minutes to plan and twenty minutes to write an essay on the topic assigned.

The instructions to the teacher included the following paragraphs:

Please encourage your students to make every effort to do well on the essay so that the study will yield valid information. Although this essay test is not part of the New Jersey College Basic Skills Placement Test, results from it will be used to make decisions about the test.

You may tell your students that this test is part of a study being conducted by the State of New Jersey in order to find the most appropriate length of time for a writing sample on a statewide test. Any information gathered from the study will be about New Jersey college students as a group, not about individual students. If you are going to grade the essays for your own purposes, you may want to inform the class of your plan.

The student's essays were scored at two special readings, one for the essays on each topic. The five readers at each reading were all college English faculty members and had all participated in previous essay-scoring sessions for the NJCBSPT. The score scale was the six-point holistic scale used for NJCBSPT essays, and the procedure for these readings was the same procedure used for regular NJCBSPT essay readings. Each reader received the following written instructions:

You will be reading essays written on a topic used previously in the New Jersey College Basic Skills Placement Test. These essays have been collected as part of a research project. They were written by both New Jersey high school and college students. You are to score them as you would score any essays written for the New Jersey test, using the standards that were established during the regular readings of the topic. The samples being used to establish the scoring standards are those that were used in the May reading--the first reading--of the topic, the reading that established the standards for all subsequent readings of the topic. Your goal as readers is to match your standards to those of the readers who scored the samples that May.

The purpose of the study will not be explained to you because knowing the purpose may influence the scoring. We want you to know, however, that the study is not being used to collect information on the performance of individual readers. When the study is completed--other readers will be performing a similar task at a later date [this phrase was changed for the second reading]--we will be happy to respond to any questions you may have about the study.

In addition, the director of each reading (the Chief Reader) received the following set of "reminders" to emphasize for the readers:

REMINDERS FOR THE CHIEF READER

Holistic Scoring

Read quickly for a total impression and score immediately.

Read supportively, rewarding for what has been done well rather than penalizing for what has been done badly or not done at all.

Compare responses. The papers are being judged in relation to each other. Use your range finders [previously scored essays] to help you make the necessary comparisons.

The six point scale

We will be using our usual six-point scale. The first decision you should make is whether the paper is upper half or lower half. Then decide where it belongs in the upper or lower half of the scale.

The total score will be the sum of two readers' scores. Do not attempt to guess what the second reader will award as the score. Give the score that you, in your best judgment, consider the paper deserves. Discrepancies will, as usual, be those scores that are more than two score points apart.

The topic

The topic has been chosen to permit the writer to respond in any number of ways, all acceptable. No paper is considered off topic unless the writer writes on another topic entirely. (Read the topic with the group and discuss the requirements of the topic with them. Mention the need to be aware that all examples are considered to be of the same worth. Mention also that responses will vary in approach and that one variation is not intrinsically better than other. That is, starting with an analysis of the task is not inherently better than starting with personal reaction to the task or a description of the task.)

The scoring

Remind readers of how to enter scores, where to find their scoring codes, and other such matters; remind them that table leaders will do quality-control checks.

Each reading began with a reading of several previously scored essays, called "range finders". The readers read and graded these papers independently. The Chief Reader then tabulated their scores for each of these essays on a large chart and told them what score each paper should have received. In some cases, the readers briefly discussed the scoring of the essay and the reasons for assigning a particular score. This procedure was repeated with another, smaller sample of previously scored essays,

until the Chief Reader was satisfied that the readers were "on scale". The readers then proceeded to score the essays written for the study.

The papers from the eight experimental groups were thoroughly mixed, to avoid any bias or dependence that might occur because of shifts in readers' standards or context effects. All essays were scored once; they were then mixed again, re-distributed and scored a second time, with each paper being scored by a different reader the second time. The readers recorded their scores on the essay booklet in code, so that no reader would inadvertently see what score another reader had assigned.

After the reading, the identifying information and the scores assigned to each essay were "scanned", i.e., electronically transferred to computer files. The two files were then "match-merged", resulting in a single record for each participating student. Any records with scores for only one essay were deleted from the file, as were those that lacked the information necessary to associate each score with a topic (G or H), a time limit, and a sequence (first or second).

The first step in analyzing the data was to check the reliability of the scoring process. Tables 2a and 2b show the joint distributions of scores assigned on first and second readings of Topics G and H, respectively. The scores assigned on first and second readings were identical for 67 percent of the Topic G essays and 57 percent of the Topic H essays. Only 1 percent of the Topic G essays and 4 percent of the Topic H essays showed a difference of more than 1 point (on a scale of 1 to 6) between first and second readings.

The correlation between scores assigned on first and second readings was .83 for Topic G and .73 for Topic H. Using the Spearman-Brown formula,

these correlations translate into reading reliability coefficients of .91 (for Topic G) and .84 (for Topic H) for the sum of scores assigned on both readings. All further analyses were done on the scores that resulted from summing the first and second readings.

Results

The results of this study revealed some complex interrelationships involving the time limit, the topic, and the ability of the students. Probably the clearest way to sort out these effects is by means of a graph such as Figure 1. This graph contains four lines, one for each of Groups 1, 2, 3, and 4. Each line is labeled with the sequence of time limits and topics presented to that group. For example, the line for Group 2, which wrote first for 20 minutes on Topic H and then for 30 minutes on Topic G, is labeled "20H, 30G". The horizontal scale represents the student's writing ability, as indicated by the student's average score on the two essays. The vertical scale represents the estimated difference between scores on the 30-minute essay and the 20-minute essay, for a typical student at a given average score level. The vertical distance between the line for each group and the zero line represents the combined effect of the difference in topics and of the extra ten minutes of writing time. This distance is clearly not the same for the four groups, and in at least two of the groups it clearly depends on the students' ability.*

*The lines were determined by linear least-squares regression. An analysis of the residuals showed no evidence of curvilinearity. The residual standard deviations for the eight groups were between 1.40 and 1.72 points.

For students of low ability, neither the extra time nor the topic appears to make much difference in their essay scores. In all four groups, the typical difference between the scores these low-ability students received under the two different time limits was about zero - no difference.

For the middle-ability and high-ability students the picture is quite different. Topic G appears to have been much easier than Topic H for these students. The good writers who received an extra ten minutes on Topic G tended to write better essays than they wrote with the shorter time limit on Topic H - better by as much as a full point on the 12-point scale. The good writers who received an extra ten minutes on Topic H tended to write essays that were slightly poorer than the essays they had written on Topic G with the shorter time limit. For students whose two scores average 10, the effect of the extra time appears to be a little less than half a point. Having the easier topic tends to raise these students' scores by almost an additional three-fourths of a point while having the harder topic tends to lower them by the same amount, so that, for these students, Topic G is nearly $1\frac{1}{2}$ points easier than Topic H..

One factor that does not seem to matter, for the good writers or the poor writers, is the sequence of topics and time limits. In each case, the two groups that received the same time limits for the two topics, but in the opposite order, performed similarly.

An analysis of variance* of the scores of Groups 1 to 4 showed the two large effects - those of the difference between topics and the interaction

*This analysis treated the student's ability, indicated by his/her average score for the two essays, as a continuous variable and assumed its effect to be linear. Factors were entered stepwise: intercept, then main effects, then two-way interactions, then the three-way interaction.

between the topic, the student's ability, and the difference in time limits - to be quite unlikely to have occurred purely by chance ($p = .0001$ and $.03$, respectively). The overall effect of the extra ten minutes was very small in relation to the other sources of variation, and the main effect of the student's ability on the difference between the 20-minute and 30-minute scores fell far short of statistical significance ($p = .23$).

Figure 2 is the same type of graph as Figure 1, but it shows the results for Groups 5 to 8. For these students, Topic G appears to be about one point easier than Topic H at all ability levels. (Note that the line for Group 5 is about one point below the line for Group 6, and the line for Group 8 is about one point below the line for Group 7.)

For the good writers, the effect of the planning period appears to be similar to the effect of the extra ten minutes that Groups 1 to 4 received. It tends to raise their scores by about half a point, on the average. For the poor writers, the planning period actually seems to result in lower scores if it comes on the first of two essays. If it comes on the second essay, it may help slightly.

An analysis of variance of the scores of Groups 5 to 8 showed a statistically significant effect for the difference between topics ($p = .0001$) and for the interactive effect of the planning period and the student's ability ($p = .03$). However, the overall effect of the planning period appeared somewhat unlikely to have occurred by purely by chance ($p = .11$).

A second phase of the data analysis focused on the students in the middle ability range - those whose average score for the two essays they wrote was at least 6 but not more than 8. These are the students whose placement is most in doubt. Table 3 shows the means and standard

deviations of the score difference variable (30-minute essay minus 20-minute essay) for these students in each of the eight groups. Table 3 also shows the mean and standard deviation of the average-score variable, as a check on the similarity of the ability of these groups.*

The results of this second phase of the analysis are consistent with those of the first phase. Students who received the extra ten minutes on Topic G tended to write essays that were better than their 20-minute essays on Topic H. Students who received the extra ten minutes on Topic H tended to write essays that were not as good as their 20-minute essays on Topic G. An analysis of variance of the scores of the middle-ability students in Groups 1 to 4 showed a statistically significant effect ($p = .005$), estimated to be about 0.8 points, for the difference between topics. No other effects even approached statistical significance. An analysis of variance of the scores of the middle-ability students in Groups 5 to 8 also showed a statistically significant effect ($p = .0001$), estimated to be about 1.1 points, for the difference between topics. Again, no other effects approached statistical significance. The estimated effect of the extra ten minutes for the middle-ability students was about one-tenth of a point in Groups 1 to 4 and one-sixth of a point in Groups 5 to 8.

*As a further check, the within-group regressions of the score-difference variable on the average-score variable were computed for these students (with average scores of 6 to 8) in each group. The regression-estimated score difference for a student with average score 7 was quite close to the mean score difference for students with average scores of 6 to 8 in each group; the differences ranged from .00 to .06 across the eight groups.

Discussion

This study was an attempt to answer the question, "What effect will an extra ten minutes of writing time or planning time have on NJCBSPT essay scores?" The results indicated that the question has no simple answer. For most of the students, the effects of the extra ten minutes will be small. None of the effects involving the difference in the time allowed was so great as to have less than a five percent probability of occurring purely by chance.

Nevertheless, the extra ten minutes did appear to have some effect on the scores of some students. For the better writers, the extra time may improve scores by an average of about half a point on the 2-to-12 scale. For these better writers, the benefit of an extra ten minutes of writing time appears to be about the same as that of a ten-minute planning period preceding the writing period.

For the poorer writers, the extra ten minutes appears to affect their scores only if it takes the form of a planning period, and the effect appears to depend on whether the students have recently had a similar writing assignment. For poor writers who have not recently written a similar exercise, the planning period may tend to result in lower scores, rather than higher scores.

For the middle-ability students - the ones for whom the placement decision is most in doubt - the extra ten minutes appears to make very little difference in their essay scores, regardless of whether it takes the form of a extended time limit or a planning period.

Table 1. Mean and standard deviation of average essay scores (both essays combined) of students in each group.

<u>Group</u>	<u>Number of students</u>	<u>Average essay scores</u>	
		<u>mean</u>	<u>standard deviation</u>
1	84	7.08	1.67
2	48	6.74	2.23
3	63	6.60	1.51
4	68	7.01	1.47
5	78	7.42	1.81
6	76	6.87	1.87
7	52	8.09	1.68
8	43	6.19	1.65

Table 2a. Scores assigned on first and second readings of Topic G

First Reading	Second Reading						Total
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	
1	7	8	1	0	0	0	16
2	2	39	17	0	2	0	60
3	0	13	97	43	1	0	154
4	0	0	18	150	33	0	201
5	0	1	0	12	42	7	62
6	0	0	0	0	9	10	19
Total	9	61	133	205	87	17	512

Table 2b. Scores assigned on first and second readings of Topic H.

First Reading	Second Reading						Total
	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	
1	6	17	2	0	0	0	25
2	3	36	25	1	0	1	66
3	0	18	118	50	3	1	190
4	0	2	39	112	19	2	174
5	0	0	2	18	18	4	42
6	0	0	0	4	9	2	15
Total	9	73	186	185	49	10	512

Table 3. Mean and standard deviation of average score and score difference (30-minute essay minus 20-minute essay) for middle-ability students in each group

<u>Students with average score 6 to 8</u>							
Group	Conditions		Number of students	<u>Average Score</u>		<u>Score difference</u>	
	1st	2nd		mean	SD	mean	SD
1	20G	30H	45	6.99	0.65	-0.16	1.40
2	20H	30G	13	7.04	0.75	0.54	1.13
3	30G	20H	41	6.83	0.72	0.49	1.57
4	30H	20G	45	7.09	0.71	-0.44	1.55
5	20G	(10+20)H	41	7.17	0.71	-0.39	1.48
6	20H	(10+20)G	43	6.97	0.72	0.63	1.53
7	(10+20)G	20H	24	7.31	0.62	0.79	1.35
8	(10+20)H	20G	25	6.92	0.64	-0.40	1.29

Figure 1: Difference in essay scores:

30-minute essay minus 20-minute essay

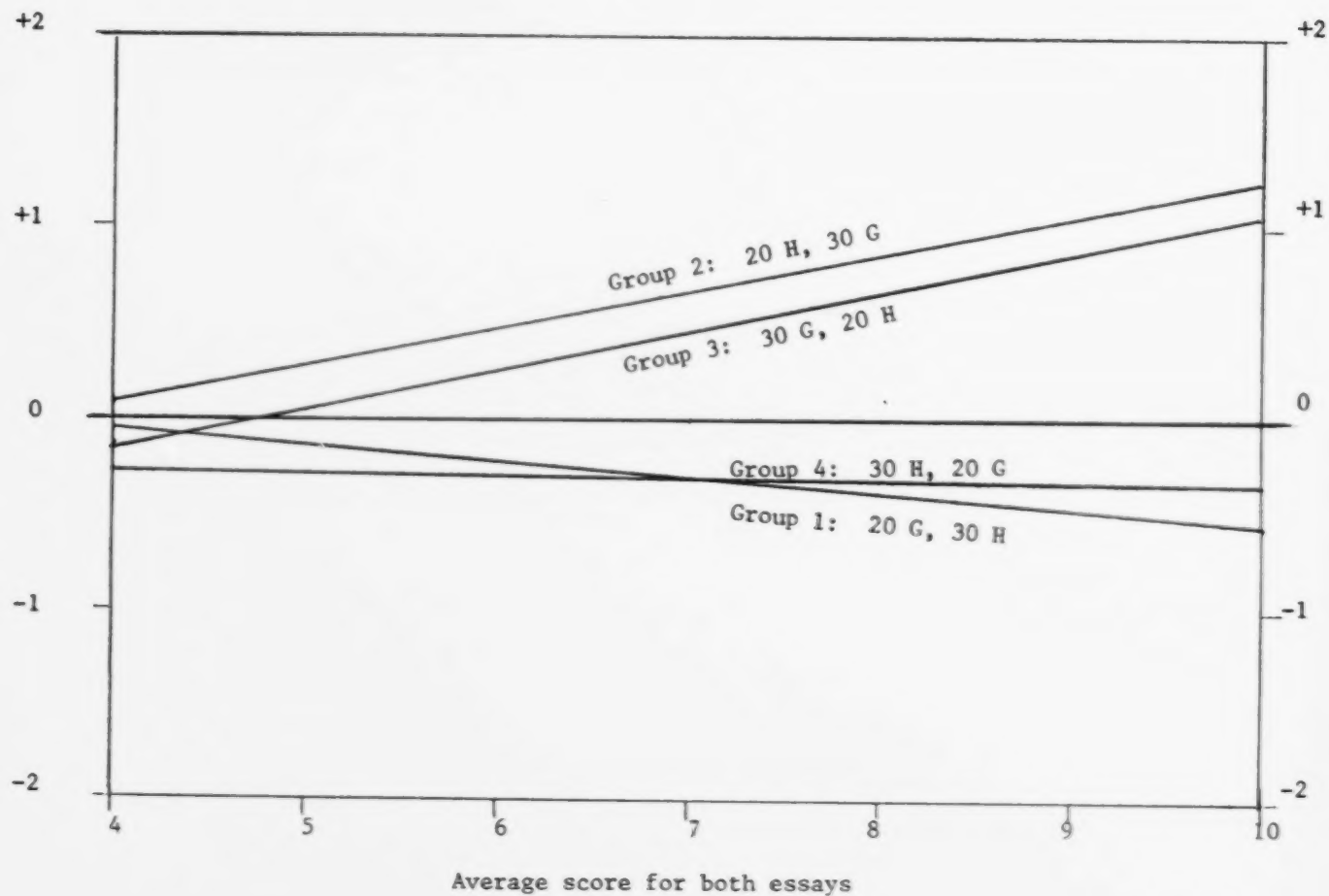


Figure 2: Difference in essay scores:

Essay with planning period minus
essay without planning period

